

Improving Courier Service Network Efficiency through Consolidations

Wenxin Zhang*, Alexia P. Payan[†] and Dimitri N. Mavris[‡]
*Aerospace Systems Design Laboratory, School of Aerospace Engineering,
Georgia Institute of Technology, Atlanta, GA, 30332-0150, USA*

Service network design is a significant consideration for courier companies because an efficient design reduces operating costs while maintaining service quality. While companies typically rely on subject-matter experts knowledge to modify their service network design on a regular basis based on changes in demand, some of them have also developed an optimization-driven approach to improve the design of their service network in the long-term. Typically, service networks are based on a hub-and-spoke design. However, operating costs may be reduced by adding consolidations on the pickup and/or the delivery routes into and out of hubs. Consolidations are locations where packages can be aggregated from multiple spokes to go into a hub or can be disaggregated to be delivered to multiple destinations from a hub. This service network design feature ultimately reduces the number of aircraft used on each route and therefore decreases the operating costs. In this study, we use Integer Programming with hierarchical objectives to generate consolidation options. The proposed algorithm accounts for network-wide demand considerations and aims at reducing costs from operating several modes of transportation by minimizing the number of consolidation locations while ensuring that every package is served and gets delivered on time at its intended destination. The algorithm is being implemented on the entire domestic U.S. market and has the flexibility to generate one or more consolidation options for each group of packages going from a given origin to a given destination. Results from the optimization are compared to solutions from a heuristic approach based on a series of geographical and operational rules. Results show that the optimization approach is able to generate better consolidation options compared to the heuristic approach. In particular, allowing packages to consolidate at a maximum of three consolidation locations results in a two percent reduction in the total costs over individual days of operations, and in nearly a one percent reduction in the total costs over a week of operations, for similar computational times. Although these reductions seem small, operating costs for courier companies tend to be in the millions or billions of dollars. Therefore, even a one percent reduction is significant.

I. Nomenclature

s	=	A spoke (origin or destination) airport
c	=	A consolidation airport
h	=	A hub airport
S	=	Set of spoke airports
C	=	Set of consolidation airports
H	=	Set of hub airports
t_{ab}^T	=	Driving time between city a and city b by truck
$A_{\#}$	=	Size of set A
x_{abc}	=	Opportunity for packages going from airport a to airport b to consolidate in airport c
$Min(F)$	=	Minimum of function F

*PhD Candidate, Aerospace Systems Design Laboratory, School of Aerospace Engineering, 270 Ferst Drive, Atlanta, GA, 30332-0150, AIAA Member

[†]Research Engineer II, Aerospace Systems Design Laboratory, School of Aerospace Engineering, 270 Ferst Drive, Atlanta, GA, 30332-0150, AIAA Member

[‡]S.P. Langley Distinguished Regents Professor, Boeing Regents Professor of Advanced Aerospace Systems Analysis, School of Aerospace Engineering, Director Aerospace Systems Design Laboratory, 270 Ferst Drive, Atlanta, GA, 30332-0150, AIAA Associate Fellow

II. Introduction

A. Courier Industry

A courier company provides commercial package delivery service to customers over small and large distances [1]. The courier industry in the U.S. emerged more than a hundred years ago, and it has seen a rapid growth in the last few decades, mostly due to globalization and the e-commerce. It is very likely that this industry will keep its momentum going into the future [2].

Currently, FedEx and United Parcel Service (UPS) are the two largest courier service providers in the world. In 2019, FedEx revenue was \$69.7 billion while UPS revenue was \$74.1 billion [3, 4]. Their revenues come from a broad range of transportation, e-commerce, and business services. FedEx revenue comes primarily from FedEx Express, which is its air division. FedEx Express possesses the world largest cargo fleet with more than 660 airplanes [5]. UPS, on the other hand, operates over 260 aircraft and relies more on its ground service [6].

Due to the capital-intensive nature of the business, a courier company faces large fixed costs [7]. In particular, the costs of operating various types and sizes of equipment [8] to transport packages make up a very large proportion of the total costs of courier companies. Therefore, in order to increase profit and gain competitive advantages, a common strategy of courier companies is to reduce equipment operating costs [9]. This can be achieved by optimizing the design of the service network and the scheduling of equipment types and sizes within this network depending on the demand for delivery services.

B. Service Network Design

The service network is the key infrastructure through which a courier company provides its delivery service [10]. The design of this network determines the connections between various locations within the network, the flows of packages throughout the network, and the movements of various types of equipment to transport packages from their origins to their destinations [11].

The design of a service network must guarantee that all packages are delivered strictly within designated time windows at their intended destinations, which is the primary requirement for courier service companies [11]. On top of satisfying delivery time constraints, the design of a service network should be as efficient as possible with regards to the number of equipment of each type used. In other words, the routes within the network should be designed such that all packages may be carried from their origins to their destinations using the smallest number of each equipment type. This means that each equipment needs to be as “full” of packages as possible. In this context, network efficiency is closely related to equipment “load factor”: the higher the load factor of the equipment used in the network, the higher the network efficiency. In turn, such an efficient service network design minimizes the costs of operating various types of equipment across the network. In a nutshell, a well-designed service network is of great importance to a courier company because it has the ability to significantly reduce operating costs while maintaining the quality of service [12].

The design of a service network usually involves a set of locations (airports) at which various activities may occur at different times (such as *origin activity*, *sorting activity*, and *destination activity*), a collection of packages that need to be transported between an origin activity and a destination activity, and a fleet of various types of vehicles (such as aircraft of different sizes and trucks) that may transport these packages [10, 11]. A given location can have multiple activities associated with it, typically occurring at different times of the day or night. Origin activities are when packages originate at a given location, destination activities are when packages arrive at a given destination, and sorting activities are when packages are sorted at a given location which is usually referred to as a “hub” location. All the packages need to be transported by a set of vehicles from their origin activities to their destination activities. The vehicles can generally be categorized into two groups: high-cost vehicles (usually aircraft) and low-cost vehicles (usually trucks). Although the origin and destination activities are always fixed for a given group of packages, there exist several routing and equipment options for this group of packages to be transported from its origin to its destination. In other words, the design of a service network consists in determining the most efficient set of routes and the most efficient equipment types to use on each route to transport all the packages within the network from their origins to their destinations [13].

C. Network Structure

Network structure is a fundamental part of a service network design. It determines how locations are connected with each other.

1. Hub-and-Spoke

Historically, most courier service providers have used a hub-and-spoke network structure. In this case, there are two types of locations, hubs and spokes, and all spokes are connected directly to one or more hubs, as shown in Figure 1 for a network with one hub.

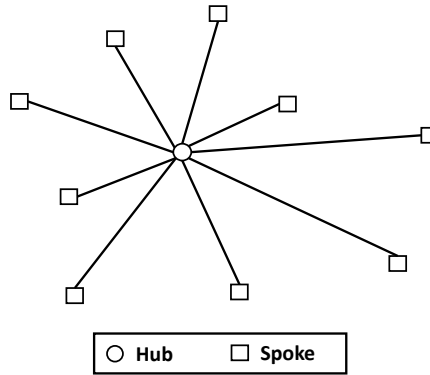


Figure 1 Hub-and-spoke structure with one hub.

In this structure, almost all the packages go through the following process. First, they go from their origins to a hub, where they get sorted based on their destinations. Then, they travel from a hub to their destinations. A small number of packages may go directly from their origin to their destination via a point-to-point route due to time constraints or other operational constraints. As a result, most packages will pass through a hub before getting to their final destination. In this case, the transportation of a group of packages that share the same origin activity and the same destination activity can be split into two phases: a pickup phase which is when the packages are picked up at their origin to be later sorted at the hub, and a delivery phase which is when the packages get delivered to their destination after having been sorted at the hub [11]. As a consequence, the route from the origin to the hub is sometimes referred to as the *pickup route*, and that from the hub to the destination is sometimes referred to as the *delivery route*.

A feature of this network structure is that each origin-hub and hub-destination route must be serviced by as many vehicles as required to transport all the packages on these routes. In order to minimize operating costs for this network structure, it is necessary to carefully assign equipment sizes to each route within the network to avoid transporting small amounts of packages on large vehicles. Nevertheless, courier service providers have limited numbers of each type of equipment in their fleet and thus the appropriately sized vehicle may not be available for each and every route in the network, leading to lower load factors and network inefficiencies as discussed previously. This is an inevitable shortcoming of the hub-and-spoke network structure [12].

2. Hub-and-spoke with Consolidations

The aforementioned shortcoming of traditional hub-and-spoke networks may be alleviated by incorporating consolidation locations into the structure [12–15]. A consolidation location acts as an intermediate stop between multiple origin spokes and a hub (on the pickup side) or a hub and multiple destination spokes (on the delivery side), as shown in Figure 2 for a network with one hub.

On the pickup side, the goal of a consolidation activity is to “aggregate”, at a single point in space, packages going from different pickup origin activities into the same hub to be sorted. This has the benefit of increasing the load factor of the equipment going into the hub from the consolidation location. On the delivery side, the goal of a consolidation activity is to “disaggregate”, at a single point in space, packages coming from the same hub where they have been sorted and going into different destination spokes. This also has the benefit of increasing the load factor of the equipment going out of the hub into the consolidation location. Therefore, a consolidation location can usually be connected to a hub via one or more high-cost vehicles (aircraft) depending on the volume of packages going into or out of the hub, and can always be connected to spokes via one or more low-cost vehicles (trucks) depending on the volume of packages originating from or going into each spoke. A consolidation activity is characterized by two important times, namely the *due time* and the *available time*. The due time is the time before which a package must arrive in order to be processed by the activity, and the available time is the time when packages become available again to be transported to the next stop. The

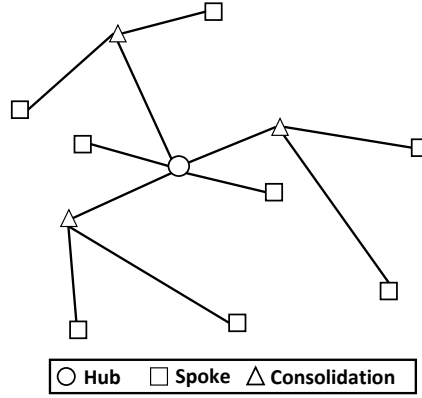


Figure 2 Hub-and-spoke structure with consolidations.

time between the due time and the available time is the *processing time* of the consolidation activity. The impact of consolidation on equipment load factors increases the efficiency of the network structure and is especially beneficial when the equipment serving the consolidation-hub or hub-consolidation leg is a higher-cost vehicle [15–17]. This does not mean that all spokes should connect to the hub via consolidation locations. In some cases, a direct connection between the spoke and the hub is more efficient as exemplified in Figure 2 with routes featuring consolidation locations and others featuring direct routes from spoke to/from hub.

In summary, the presence of consolidation locations improves the efficiency of a hub-and-spoke network structure by increasing the load factors of vehicles (especially high-cost ones) used overall within the network.

III. Literature Review on the Use of Consolidations in Service Network Design

Service network design problems have been extensively studied. Although most past work focus on the optimization of the hub-and-spoke network structure, some work consider the addition of consolidation locations to the traditional hub-and-spoke network or propose similar strategies to increase network efficiency and reduce operating costs.

For example, Kuby and Gray [12] worked on improving the traditional hub-and-spoke network, which was developed by O’Kelly [18–20], by incorporating stopovers and small aircraft, called feeders. They assumed that there is a single hub which location is predetermined, that daily package volumes and aircraft schedules are fixed day-by-day, and that only packages that need to be delivered overnight are considered. They used a mixed-integer optimization approach to find the optimal aircraft routes. In order to generate potential paths from origin to destination, they applied four criteria: time window, maximum number of stopover airports (which is two in their case), geographical rule, and aircraft size limitation. They applied their theory on a FedEx data set for the western U.S. market, and compared the efficiency of the traditional hub-and-spoke network with the same network with stopovers and feeders added. They concluded that the stopovers and feeders can significantly reduce costs, flight distances and the number of aircraft used in the hub network. They also claimed that a network with only direct flights from hubs to spokes is not realistic. The stopover and feeder concept that they use is not exactly the same as consolidation, in that a stopover flight uses the same airplane along the route between a hub and a spoke, while a consolidation usually has two different types of vehicles going in and out (a truck and an airplane). Yet, their work is another valuable solution to improve the traditional hub-and-spoke network.

Ruckle [21] created a Collapsed Product Set Sequential Algorithm to solve the network design problem. She tackled the consolidation problem in an “a posteriori” manner. That is, she considered consolidation opportunities after all hub-spoke aircraft routes were determined, and she used a heuristic approach based on rules to generate consolidation opportunities. Although the “a posteriori” method makes it easier to identify good consolidation opportunities compared to an “a priori” method in which aircraft routes are not determined, it also reduces the flexibility of the algorithm, which would potentially eliminate some good consolidation opportunities. The use case featured a data set spanning an entire week of operations and including the entire U.S. domestic market. She concluded that the proposed Collapsed Product Set Sequential Algorithm works very well without consolidations, but poorly with consolidations, because results showed that including consolidations dramatically increased the runtime without improving much the solution quality.

Finally, Patel [22] worked specifically on generating consolidation options for a network using an unsupervised

learning approach. Instead of using heuristic methods, he used the K-Mean Clustering technique to generate consolidation options. He applied the proposed approach on a data set within the California region, assuming no incoming flights and the aircraft leaving the region all going to a regional hub, which is a relatively small scale problem. In his K-Mean algorithm, he tried to cluster cities that are geographically close to each other to create consolidation options and set a constraint that all clusters should be able to transport the entire package load within the California region. However, it would have been more appropriate to consider the volume at each consolidation, because that would have allowed the determination of the load factor of airplanes, which is a more direct indicator of network efficiency. He concluded that the approach resembles shipping movements in the California region, without showing the actual impact on metrics of interests, such as costs and number of flight hours.

The aforementioned studies have made significant contributions to the network design problem with consolidations. However, the approaches proposed to generate consolidation options were mostly heuristic, based on a set of problem-specific rules, and do not account for *network effects*. Network effects refer to the concept that any consolidation decision made in the network will not only affect the volume of packages at that location, but also the volume of packages at other consolidation locations. Due to this effect, rule-based heuristic methods will eliminate potentially good consolidation options, reduce the flexibility of the algorithm, and prevent the optimization solver from finding better solutions. The main goal of this study is to find a better approach to the generation of consolidation options, so that it takes network effects into account and thus improves the potential of the solver to find better solutions.

An *optimization* is the maximization/minimization of an objective function relative to some variables, often representing a range of choices under some constraints. When an optimization has more than one objective function, the problem becomes a multi-objective optimization problem [23, 24]. One way to solve this type of problems is hierarchically by assigning a priority to each objective function and optimizing for the objectives in decreasing order of priority. At each step, the algorithm finds the best solution for the current objective, but only from solutions that would not degrade the solution quality of higher-priority objectives (or degrade their solution quality within a certain level). This way of solving an optimization problem with multiple objectives is called hierarchical multi-objective optimization [25]. When an optimization problem has its variables as integers, it becomes an integer programming (IP) optimization problem [22, 25].

Compared to heuristic methods, an optimization-based approach has the flexibility to incorporate network effects if the correct variables, objectives and constraints are implemented. In this paper, we propose the use of a hierarchical multi-objective Integer Programming optimization approach for creating consolidation options within a larger optimization framework to solve the service network design problem. The solution from this approach will be compared and contrasted with that of a heuristic approach to quantify the improvements obtained by using an IP optimization approach.

IV. Overall Framework

This study is part of a larger endeavor meant to solve the service network design problem and for which an optimization-driven framework has been developed. *

The workflow of the overall framework is composed of the following main steps:

- First, create a large set of routing options for packages and flight options according to the input data set.
- Second, construct an optimization problem based on the options created in the first step.
- Finally, use the commercial optimization solver Gurobi to solve the optimization problem and obtain the set of variables that minimizes operating costs. The optimal solution contains the set of routes used to transport packages, and the type of equipment used on each route.

The creation of routing options is composed of three elements: 1). the creation of *sort association* options which identifies which spoke locations are allowed to connect to which hub; 2). the creation of consolidation options which adds feasible consolidations for each spoke-hub connection; 3). the creation of two-legged flights within the network. A two-legged flight is different from a route involving a consolidation in that the flight that starts from an origin, picks up some packages at an intermediate stop and then goes into a hub; or starts from a hub, picks up some packages at an intermediate stop and goes into a destination. The equipment used in a two-legged flight is an aircraft all the way, while a consolidation route may feature two different equipment types (but always features at least a truck). Figure 3 shows the workflow of the overall optimization-driven framework [21].

Generally speaking, the overall framework converts a courier company real-life problem into a mathematical optimization problem and then solves it by using the Gurobi solver.

*The overall optimization-driven approach is different from the optimization approach used to tackle the consolidation problem in this study. In other words, the proposed optimization method for solving the consolidation problem is “an optimization within a larger optimization framework.”

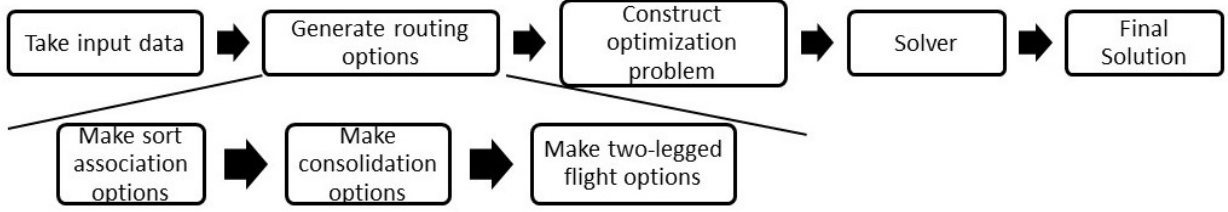


Figure 3 Overall optimization-driven framework.

In this process, the generation of flight opportunities for aircraft and routing options for packages is the most crucial step. A good mathematical solver by itself cannot yield a good solution without a properly formulated model of the problem. In this case, options need to be generated in such a way that the solver can find the optimum solution to the problem within a reasonable amount of computing time. Therefore, there needs to be a balance between the number of options generated and the time required to solve the problem: too many options will result in an unacceptable runtime, but too few options might result in the inability of the solver to find a good solution to the problem. In this case, it is necessary to generate only feasible and realistic/relevant options. For example, for packages originating in Los Angeles and going to Las Vegas, a routing option involving a hub on the east coast is not the most efficient way to route the packages and probably not realistic operationally. On the contrary, a routing option featuring a hub located on the west coast is much more realistic. Given the size of the problem (featuring several types of packages, hundreds of spokes, several hubs, and several equipment types and sizes), generating routing options that are both feasible and relevant operationally will enable the solver to find a better solution to the problem faster.

This study focuses on the generation of feasible and relevant consolidation options (involving the determination of both the number of consolidation activities and their locations) in order to reduce flight hours and operating costs (and additionally driving hours, also called truck hours), while still transporting the maximum amount of packages in the network. We propose a hierarchical multi-objective integer programming optimization approach and implement it on a large dataset representative of a courier service company. We then compare the solution obtained from the optimization-based approach with that obtained by a heuristic approach to determine improvements in metrics such as number of flight hours and objective function value representative of operating costs.

V. Heuristic Consolidation Approach

The heuristic approach is based on problem-specific rules and filters. The creation of consolidation options builds upon the creation of sort association options, i.e. allowed hub-and-spoke connections. In practice, adding consolidation locations within the hub-and-spoke connections is essentially making consolidation options for hub-spoke pairs.

The heuristic approach starts with the creation of sort association options. Then, all time feasible consolidation options are created. The concept of time-feasibility means that a consolidation option should guarantee that packages have enough time to go from their origin spoke at the time they become available, to the consolidation location, and then to the hub for the pickup side (or vice versa to the destination spoke for the delivery side) before its due time at the hub (or the destination spoke for the delivery side).

Next, time-feasible consolidation options are trimmed down using a set of rules, as follows:

- Sort Connection Rule: a spoke and its consolidation location should be allowed to connect to the same hub
- Volume Rule: origin spokes with smaller volumes of originating packages go to consolidation locations with larger volumes of originating packages (pickup side), or consolidation locations with larger volumes of originating packages go to spokes with smaller volumes of arriving packages (delivery side)
- Truck Rule: spokes do not use consolidation locations if the packages originating from or arriving at the spokes can be transported via a truck to the hub
- Geographical rule: no backtracking allowed in spoke-consolidation-hub or hub-consolidation-spoke route), see the aforementioned Los Angeles - Las Vegas example
- Parking Rule: consolidation locations must have at least two aircraft parking spots

At this point, if a hub-spoke pair still has more than one consolidation option, the final consolidation location is selected according to one of the following filters: it is closest to the spoke, or it has the most aircraft parking spots (i.e. it is located at the largest airport). Figure 4 shows the steps of the heuristic consolidation approach.



Figure 4 Flowchart for the heuristic consolidation approach.

With this heuristic approach, each hub-spoke pair has at most one consolidation option: some pairs do not have any consolidation option. Besides, there is no criterion to check the quality of the consolidation options generated with this approach since it is based on problem-specific rules and filters.

Furthermore, the heuristic approach does not take into account the network effects. The service network considered consists of large numbers of locations, packages, and vehicles which are interacting with each other. Thus, the routing options are tangled, meaning that when we add or remove an option, other options will be affected. This effect is a fundamental feature of a network which is not accounted for in the heuristic approach.

In this study, we propose a more robust and less problem-dependent optimization-based approach to create feasible and relevant consolidation options that will overcome the major disadvantages of the heuristic approach and will account for network effects.

VI. Optimization-Based Consolidation Approach

In this study, we consider an initial set of consolidation options, which need to be down-selected. Since an option will be either selected or not, all the options can be considered as decision variables in an integer programming problem.

In this section, we will first introduce the single-run optimization which selects the best consolidation option for each hub-spoke pair from all time-feasible candidates. Then, we will introduce the multiple-run optimization which generates a pre-specified number of “best” consolidation options for each hub-spoke pair through an iterative process.

As mentioned in the previous section, the starting point of this study is a set of options for spoke-hub connections. Then, we generate all time-feasible consolidation options for each hub-spoke pair in the network of interest. The goal of this study is to downselect “good” consolidation options from this initial set of time-feasible candidates for each spoke-hub connection using an integer program (IP) with hierarchical objectives.

The single-run optimization formulation is composed of three main elements.

- Decision variables. All time-feasible options are converted into decision variables, i.e. for each consolidation option x_{shc} , its value can be either 1, indicating it is selected by the optimizer, or 0, indicating the opposite.
- Objective functions. The primary objective is to minimize the number of consolidation locations because fewer consolidation locations lead to fewer aircraft used in the network (a consolidation uses aircraft to connect to a hub), and the cost of operating aircraft in the network is the primary concern in this problem. The secondary objective is to minimize the travel time associated with transporting packages using trucks between spokes and consolidation locations.
- Constraint. Each spoke-hub pair must have only one consolidation option, so that no pair is left without any. Due to this constraint, each pair ends up with exactly one consolidation option.

The optimization problem is therefore formulated as follows:

- Binary variables — consolidation option for a given spoke-hub pair:

$$x_{shc} \quad (1)$$

- Objective functions:

Primary objective — minimize the number of consolidations in the network:

$$\text{Min}(\mathbf{C}^\#) \quad (2)$$

Secondary objective — minimize the travel time by truck between spokes and consolidation locations:

$$\text{Min}(\sum_{s \in \mathbf{S}} \sum_{h \in \mathbf{H}} \sum_{c \in \mathbf{C}} t_{sc}^T x_{shc}) \quad (3)$$

- Constraint — each spoke-hub pair has exactly one consolidation option:

$$\sum_{c \in \mathbf{C}} x_{shc} = 1 \quad (4)$$

We finally use the mathematical optimization solver Gurobi[†] to solve the aforementioned optimization problem.

The single-run optimization returns a single consolidation option for each spoke-hub pair. However, the goal of this study is to generate a number of “good” consolidation options for each spoke-hub pair that the optimizer can choose from. This is achieved by running the previous algorithm multiple times.

In the multiple-run optimization, the “best” consolidation option for each spoke-hub pairs obtained from the single-run optimization is removed from the set of time-feasible consolidation options and the algorithm is run again on the resulting reduced set of options to obtain the “second best” consolidation option for each spoke-hub pair. The number of iterations is designed to be flexible, so that a pre-specified number or less than this number of consolidation options may be generated for each spoke-hub pair. In each individual run, only the input set changes. The optimization algorithm stays the same. After the specified number of runs have been performed, the solutions obtained from each run are aggregated into a final set of consolidation options.

Figure 5 summarizes the process followed by the multiple-run optimization.

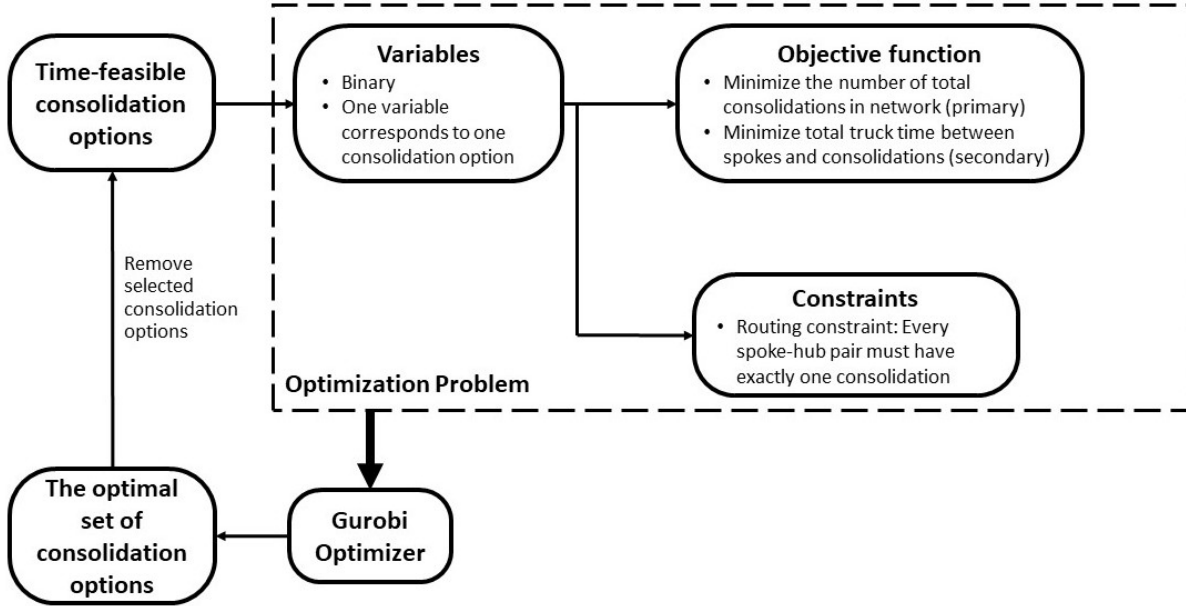


Figure 5 Multiple-run optimization process.

Figure 6 shows the entire optimization-based process to generate consolidation options: it takes as input all hub-spoke connections built beforehand, generates all time-feasible consolidation options for each spoke-hub pairs, and outputs a certain number of consolidation options for each hub-spoke connection using an optimization approach.

[†]<https://www.gurobi.com/>

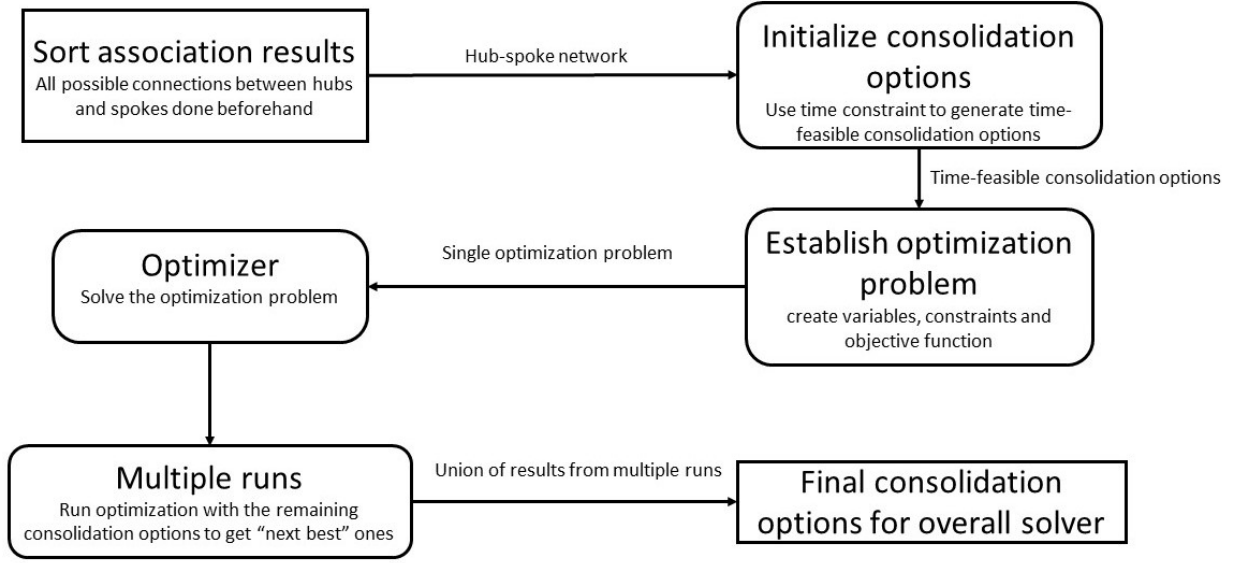


Figure 6 Flowchart for the entire optimization-based process to generate consolidation options.

VII. Implementation and Results

A. Use Case Description

Both the heuristic approach and the IP approach have been implemented to generate consolidation options within the service network of a typical courier company. In this section, we compare the results generated in each case and demonstrate improvements obtained from the IP approach.

The data set consists of volumes of packages with their origins and their destinations over an entire week of operations divided into five parts, also called “days” in this study: one for each day from Monday through Thursday, and one for the remaining three days (Friday to Sunday). The scope of the problem stays within the contiguous United States (CONUS) consisting of 48 states. We consider a total of five hubs, enumerated as Hub 1 to Hub 5, and all other airports are defined as spokes, whose connections to hubs are obtained from the sort association process.

A sort is an activity that happens at a hub, where incoming packages are sorted based on their destinations. Similarly to a consolidation activity, a sort activity has a due time and an available time. An individual day is split into two parts, namely day and night, for the activities happening at a hub. The sorting activity happening during the day at a hub is called a “day sort”, while the sorting activity happening during the night at a hub is called a “night sort”. Consequently, there are ten different sort activities in the data set considered.

Using the aforementioned data, two sets of consolidation options are generated:

- One for each individual day mentioned earlier in order to get the package network design and schedule for each individual day, called “individual-day options.”
- One in which the individual-day options are combined into a set of options for an entire week of operations, called “week options.”

Then, the heuristic approach is applied to each individual day to obtain baseline consolidation options and baseline package network and flight schedules within the overall optimization-driven framework. The IP approach is also applied to each individual day three times to generate one, two, and three consolidation options at most for each spoke-hub pair in the network. We also consider a reference day (the one that corresponds to the largest volume of packages being transported within the network) and use the IP approach to generate four, five, and six consolidation options at most for each spoke-hub pair in the network. We use this reference day to determine the optimal number of consolidation options to generate, at most, for each spoke-hub pair in the network as will be discussed in section VII.C. Finally, both the heuristic approach and the IP approach are applied to the aforementioned data set over an entire week of operations.

In the remainder of this section, we are going to: (1) compare the consolidation options generated by the heuristic and the IP approaches to determine the effectiveness of the IP approach in general; (2) compare the runtimes and the

final solutions obtained by running the overall optimization-driven framework using the consolidation options generated by the heuristic and the IP approaches.

B. Comparison of Consolidation Options Between Heuristic Approach and Optimization Approach

Among the five hubs considered, Hub 1 has the busiest traffic, and a day sort generally handles more volume of packages than a night sort. As a result, we focus on the consolidation options generated between the various spokes in the network and Hub 1 day sort (noted HUB1/D1) in the rest of this section. As a matter of fact, similar results are obtained for all the other sort activities.

Figure 7 and Figure 8 show the consolidation options generated for HUB1/D1 generated by the heuristic approach and by the IP approach with one consolidation option for each spoke-hub pair respectively.

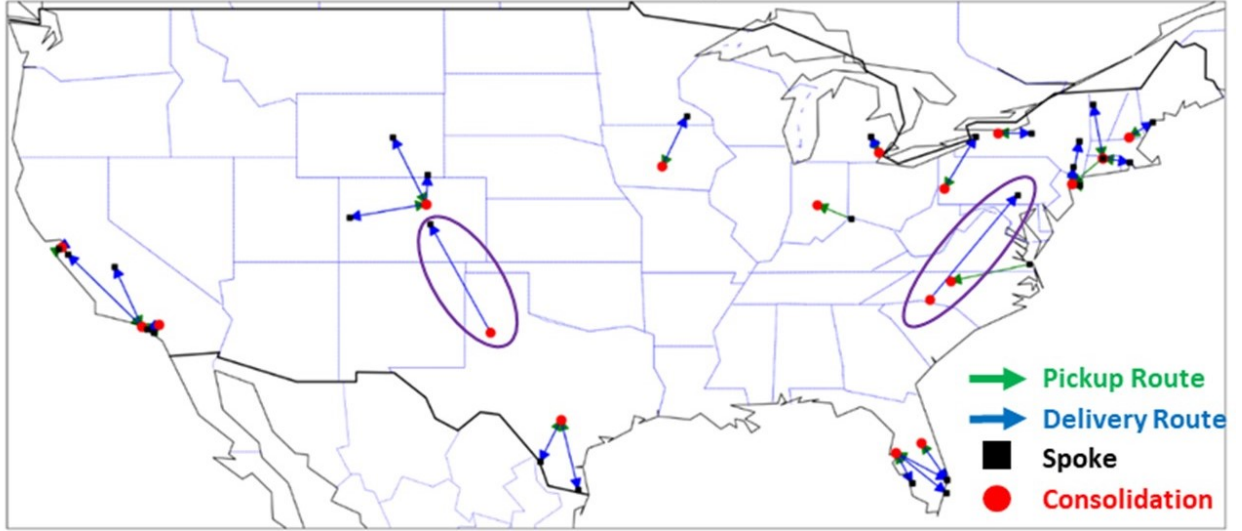


Figure 7 Consolidation options for HUB1/D1 generated by the heuristic approach.

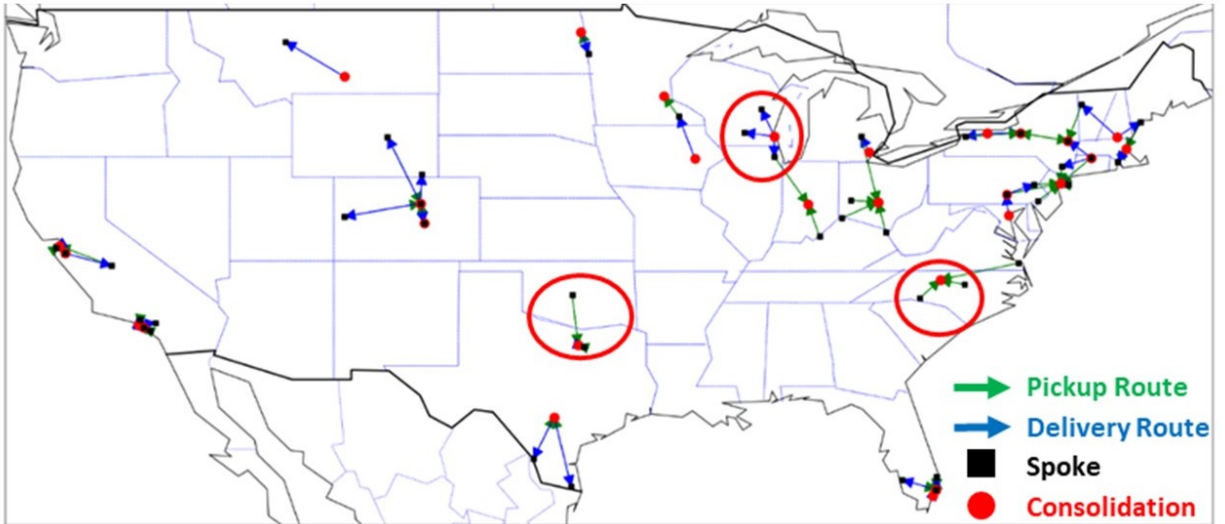


Figure 8 Consolidation options for HUB1/D1 generated by the IP approach with one option for each spoke-hub pair.

From Figure 7 and Figure 8, we observe the following:

- 1) The IP approach guarantees that every spoke-hub pair has at least one consolidation option, whereas the heuristic approach does not create any option for some pairs. Several examples are circled in red in Figure 8. For instance, the heuristic approach does not create any consolidation options for the three spoke cities in Wisconsin and Illinois, because none of the time-feasible consolidation options for these spoke-hub pairs satisfied the rules and filter. However, the IP approach does generate consolidation options for the corresponding spoke-hub pairs.
- 2) The IP approach generates “better” options in that the distance between the spoke and the consolidation location is reduced compared to the results from the heuristic approach. Examples are circled in purple ellipses in Figure 7. For instance, the heuristic approach generate a consolidation option in Texas for a spoke city in Colorado, which is probably not the most appropriate operationally. However, the IP approach generates a consolidation option within Colorado, which is much closer to the spoke city and more realistic. This improvement in the location of the consolidation activity allows reductions in travel costs via truck and truck hours, and more importantly allows spokes that are close to each other to consolidate at the same location.

Furthermore, results shows that by using the IP approach instead of the heuristic approach, the total volume of packages consolidated and involving HUB1/D1 increased by 128% on the pickup side and by 97% on the delivery side, while the average volume of packages consolidated and involving HUB1/D1 increased by 115% on the pickup side and by 50% on the delivery side. This percentage growth in volumes of packages consolidated both on the pickup side and on the delivery side indicates that the IP approach has significantly increased the total volume aggregated at all the consolidation locations, as well as the average consolidated volume, which results in higher aircraft load factors for aircraft travelling from consolidation locations to hubs and vice versa. This is because network effects are considered in the IP approach and the primary objective function is to minimize the number of consolidation locations, so that the average consolidated volume is maximized. Additionally, the number of origins/destinations that are allowed to consolidate increases when using the IP approach compared to the heuristic approach. This is because the routing constraint imposes each spoke-hub pair to have at least one consolidation option, whereas the heuristic approach does not.

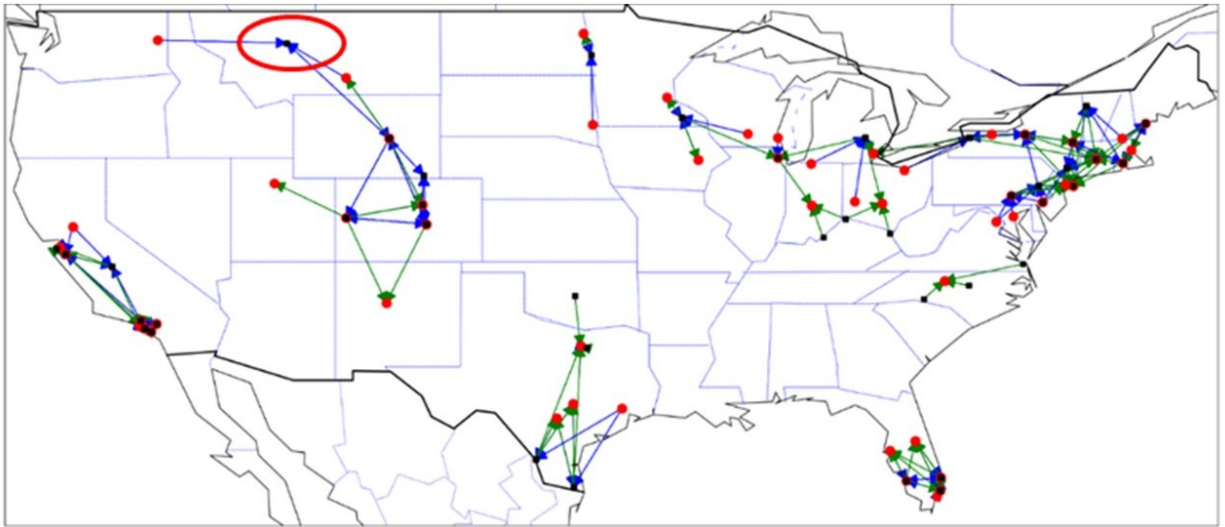


Figure 9 Consolidation options for HUB1/D1 generated by the IP approach with at most three options for each spoke-hub pair.

Figure 9 shows the consolidation options generated by the IP approach applied three times. We can see that spoke-hub pairs have at most three consolidation options. So far, we have focused on the HUB1/D1 sort activity and have analyzed the time-feasible consolidation options generated for the spokes connected to this activity. However, similar results would be obtained for the other nine sort activities and are not included in this paper for brevity. From the aforementioned results and observations, we can conclude that by using the IP approach rather than the heuristic approach, the average volume at consolidations is increased so that airplanes operating between consolidations and hubs have potentially higher load factors, which will improve the network efficiency and reduce costs.

C. Comparison of Results at the Network Level

Looking at the consolidation options alone is not sufficient to demonstrate the effectiveness of the IP approach. Therefore, we include the consolidation options generated by the heuristic approach and the IP approach into the overall optimization-driven framework and use the Gurobi solver to obtain the optimal package network and flight schedules (both for individual days and for the entire week) in each case. In this context, we focus on two comparison criteria:

- 1) The cost-related metrics, such as total operating cost, the number of hours flown by large aircraft in the network also called trunk hours, and the number of large aircraft used also called trunk hulls.
- 2) The runtime of the Gurobi solver for the overall optimization.

Due to proprietary reasons, we only display percentage changes instead of absolute values for the cost-related metrics in all the following tables. Although the percentage values in the tables may seem small, the underlying numbers are quite large (especially for the objective function which is a measure of total operating costs) so that the reductions in the aforementioned metrics are still significant.

If the IP approach is effective, its associated operating costs (objective function value) should be lower than those for the heuristics approach. Additionally, the corresponding runtime should stay comparable to the heuristic approach. If these two goals are achieved, we can be confident that proposed optimization-based IP approach outperforms the heuristic approach.

As mentioned earlier, the week is divided into five “days”: Monday to Thursday as four individual days, and Friday to Sunday as one day. The reference day has the highest throughput in a week and the other days are named Day 1 to Day 4 in the following figures and tables.

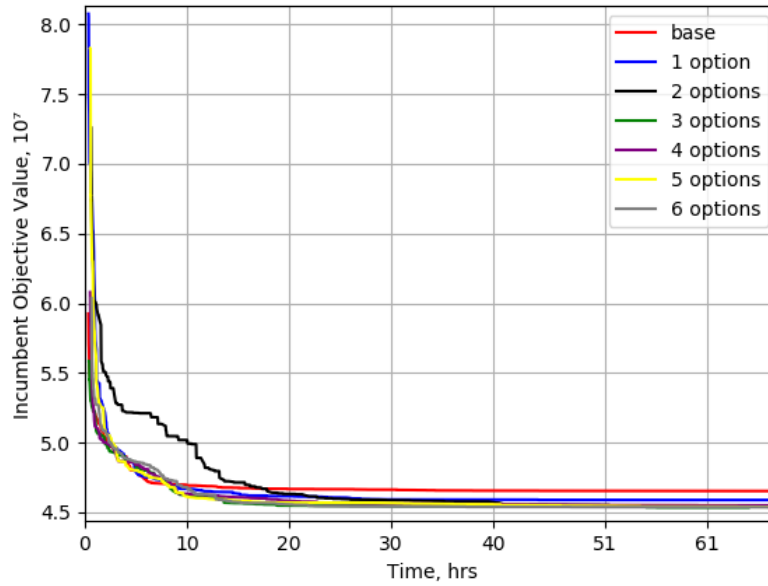


Figure 10 Objective function value change over runtime for the reference day.

Figure 10 shows the change in the incumbent objective function value, which is an indicator of total operating costs, over the runtime period. The seven colored lines correspond to different ways of generating consolidation options. The base case uses the heuristic approach, whereas the other cases use the IP approach with different numbers of consolidation options for each hub-spoke pair (one to six). For each case, the incumbent objective value starts at a large value and gradually decreases in a step-wise manner every time the solver/optimizer finds a better solution to the problem. Eventually, the objective function value levels off, meaning that the optimizer has found the optimum integer solution to the problem, and the case is considered “converged” at this point. From Figure 10, we observe the following:

- The base case ends up with the largest total operating costs, meaning that the IP approach, regardless of the

number of options generated at most for each spoke-hub pair, results in smaller objective function values, i.e. total operating costs.

- By the 48th hour, all cases have converged. This implies that the IP approach, regardless of the number of options generated at most for each spoke-hub pair, has a similar runtime to the heuristic approach.

Table 1 Comparison of Solutions for the Reference Day

	Final objective value	Trunk hours	Trunk hulls
Baseline	-	-	-
1 option	-1.41%	-1.90%	-1.31%
2 options	-2.25%	-3.98%	-1.96%
3 options	-2.56%	-4.05%	-2.29%
4 options	-2.23%	-3.28%	-1.96%
5 options	-2.42%	-4.79%	-1.63%
6 options	-2.51%	-4.24%	-1.96%

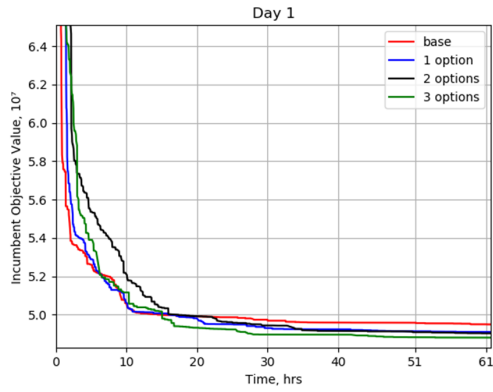
Table 1 shows that the IP approach results in lower objective function values (i.e. total operating costs), and lower numbers of trunk hours and trunk hulls over a day of operation compared to the heuristic approach. Table 1 further shows that as the number of consolidation options generated by the IP approach increases from one to three (at most), the percentage reduction increases. As more consolidation options are being generated (four to six), improvements in the objective function value, the number of trunk hours, and the number of trunk hulls do not increase anymore and tend to stagnate or worsen. Therefore, we conclude that at most three consolidation options per spoke-hub pair is optimum for the reference day.

As mentioned earlier, we implemented both the heuristic approach and the IP approach with one, two, and three consolidation options generated on the other “days” of the week as well. The corresponding results are shown in Figure 11 and Table 2. Figure 11 displays similar results to the reference day in terms of runtime and objective function value changes over time, and Table 2 shows similar results to the reference day in terms of objective function value for the optimum solution obtained, number of trunk hours, and number of trunk hulls. In summary, the IP approach always results in lower total operating costs compared to the heuristic approach without increasing much the runtime, and the lowest optimum total operating cost is achieved when the number of consolidation options generated in three per spoke-hub pair (at most).

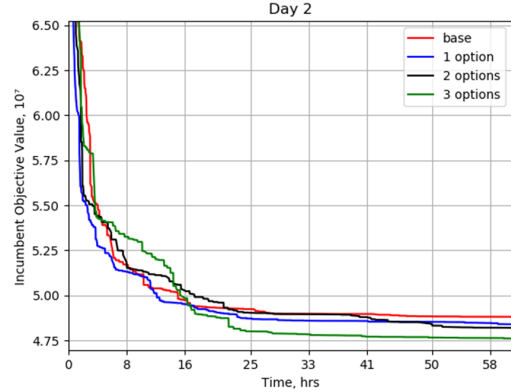
Table 2 Comparison of Solutions for the Other Individual Days

	Day 1			Day 2		
	Objective value	Trunk hours	Trunk hulls	Objective value	Trunk hours	Trunk hulls
Baseline	-	-	-	-	-	-
1 option	-0.80%	-0.72%	-0.96%	-0.56%	-1.46%	-0.32%
2 options	-0.94%	-0.79%	-0.96%	-0.98%	-1.71%	-0.97%
3 options	-1.40%	-2.80%	-0.96%	-2.16%	-3.07%	-1.94%
	Day 3			Day 4		
	Objective value	Trunk hours	Trunk hulls	Objective value	Trunk hours	Trunk hulls
Baseline	-	-	-	-	-	-
1 option	-1.19%	-2.42%	-0.66%	-0.51%	-0.67%	-0.00%
2 options	-2.34%	-4.55%	-1.97%	-2.13%	-2.40%	-1.20%
3 options	-2.59%	-3.87%	-1.97%	-2.40%	-3.34%	-1.20%

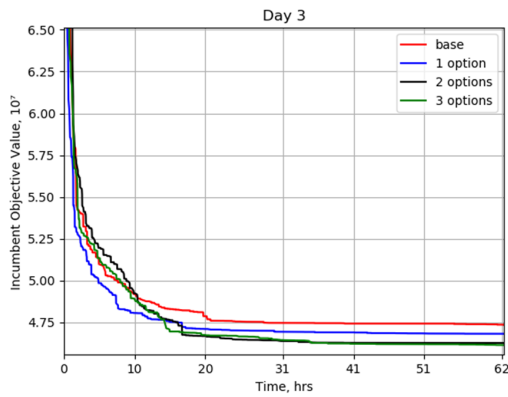
Consequently, we conclude that the IP approach is consistently better than the heuristic approach on every individual day in a week, and that at most three consolidation options per spoke-hub pair is the optimal number of consolidation options to be generated by the IP approach for the problem considered on an individual-day basis.



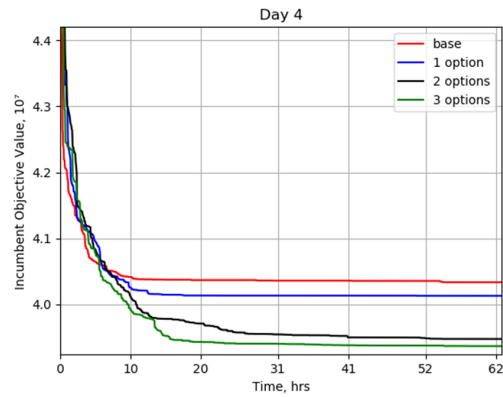
(a) Objective value change over runtime for Day 1.



(b) Objective value change over runtime for Day 2.



(c) Objective value change over runtime for Day 3.



(d) Objective value change over runtime for Day 4.

Figure 11 Objective value change over runtime for individual days other than the reference day.

Finally, we compare the results from the heuristic and the IP approach on an entire week of operations. We implement the IP approach with three consolidation options at most for each spoke-hub pair, as this was found to be the optimal number of options on an individual-day basis.

Figure 12 shows that the objective function value in both cases converges approximately after 48 hours and that the IP approach leads to a better optimum solution (lower objective function value).

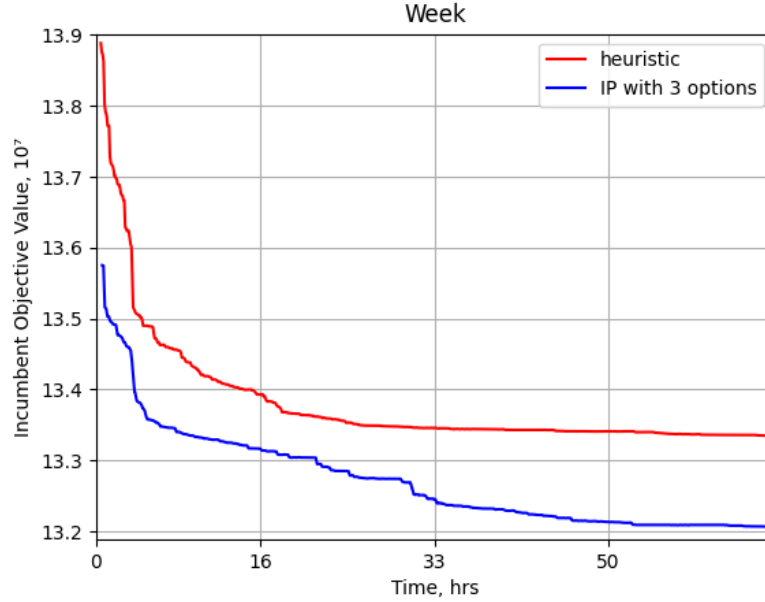


Figure 12 Objective function value change over runtime for an entire week of operations.

The statistics in Table 3 show that by using the IP approach, total operating costs, flight hours, and fleet numbers over a week of operations are significantly reduced compared to the heuristic approach.

Table 3 Comparison of Solutions for a Week of Operations

	Final objective value	Trunk hours	Trunk hulls
Heuristic	-	-	-
IP	-0.98%	-2.05%	-0.60%

VIII. Conclusions

This study showed that service network efficiency can be significantly improved by incorporating consolidations and that the proposed Integer Programming approach provides better consolidation options than a heuristic approach based on rules and filters. We applied both the heuristic approach and the IP approach to the operations of a typical courier company. By comparing the consolidation options generated by both approaches, we conclude that the IP approach increases the average volume of packages consolidated, so that the airplanes used between consolidations and hubs have higher load factors, which ultimately improves the efficiency of the network and reduces operating costs. By comparing the final solutions obtained from the overall optimizer, we conclude that the IP approach is able to successfully and consistently lower the objective function value by reducing flight hours and hull counts while keeping the runtime acceptable over individual days of operations and throughout a week of operations. Generating at most three consolidation options per spoke-hub pair using the IP approach, we showed that the IP approach is able to achieve around two percent reductions in the objective function values for individual days, and nearly one percent reduction of the total objective value over an entire week for similar runtimes compared to the heuristic approach. These are considerable improvements given the large numbers associated with the objective function for typical courier companies. This demonstrates the effectiveness of the optimization-based IP approach to generate feasible and realistic/operationally relevant consolidation options that eventually reduce the overall operating costs associated with the package network and flight schedule of a typical courier service provider.

References

- [1] “Postal and Courier Service,” World Trade Organization, 1998.
- [2] Schmidt, A., “A Look at the Courier Service Industry in the United States,” Market Realist, 2015.
- [3] Trefis, “FDX Revenues: How Does FedEx Make Money?” , 2020. URL <https://dashboards.trefis.com/no-login-required/m1AnGuQP/FDX-Revenues-How-Does-FedEx-Make-Money-?fromforbesandarticle=fdx190922>.
- [4] Macrotrends, “UPS Net Income 2006-2020,” , 2020. URL <https://www.macrotrends.net/stocks/charts/UPS/ups/net-income>.
- [5] FedEx, “FedEx Corporation Q2 Fiscal 2020 Statistics,” , Dec. 2019.
- [6] UPS, “UPS Aircraft Fleet Fact Sheet,” , Jan. 2020.
- [7] Schmidt, A., “What Are FedEx’s Major Costs?” Market Realist, 2015.
- [8] “Market share of the leading couriers and local delivery providers in the United States in 2017,” Statistica, 2017.
- [9] Porter, E., *Competitive Strategy: Techniques for analysis industries and competitors*, 1st ed., Free Press, 1998.
- [10] Barnhart, C., and Schneur, R., “Air Network Design for Express Shipment Service,” *Operations Research*, Vol. 44, No. 6, 1996, pp. 852–863.
- [11] Barnhart, C., and Krishnan, N., “Network Design for Express Shipment Delivery,” *Computational Optimization and Applications*, Vol. 21, 2002, pp. 239–262.
- [12] Kuby, M., “The Hub Network Design Problem with Stopovers and Feeders: The Case of Federal Express,” *Pergamon Press Ltd.*, 1991.
- [13] Kim, D., “Large Scale Transportation Service Network Design: Models, Algorithms and Applications,” Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1997.
- [14] Armacost, A. P., and Barnhart, C., “Composite Variable Formulations for Express Shipment Service Network Design,” *Transportation Science*, Vol. 36, No. 1, 2002, pp. 1–20.
- [15] Armacost, A. P., and Barnhart, C., “UPS Optimizes Its Air Network,” *Interfaces*, Vol. 34, No. 1, 2004, pp. 15–25.
- [16] Hane, C. A., Barnhart, C., and Johnson, E. L., “The fleet assignment problem: solving a large-scale integer program,” *Mathematical Programming*, 1995, pp. 211–232.
- [17] Louwerse, I., Mijnders, J., Meuffels, I., Huisman, D., and Fleuren, H., “Scheduling movements in the network of an express service provider,” *Flexible Services and Manufacturing*, Vol. 26, 2014, p. 565–584.
- [18] O’kelly, M. E., “The location of interacting hub facilities,” *Transportation Science*, Vol. 20, 1986, pp. 92–106.
- [19] O’kelly, M. E., “A quadratic integer program for the location of interacting hub facilities,” *European journal of operational research*, Vol. 32, 1987, pp. 393–404.
- [20] O’kelly, M. E., “Activity levels at hub facilities in interacting networks,” *Geographical Analysis*, Vol. 18, 1986, pp. 343–356.
- [21] Ruckle, L., “Approaches to Solving the Express Shipment Service Network Design Problem,” Ph.D. thesis, Georgia Institute of Technology, 2018.
- [22] Patel, S., “Improvement of Transshipment Optimization Through an Unsupervised Learning Approach,” Tech. rep., Aerospace System Design Laboratory at Georgia Tech, 2017.
- [23] Savelsbergh, M. W. P., “Local search in routing problems with time windows,” *Annals of Operations Research*, Vol. 4, 1985, pp. 285–305.
- [24] Luenberger, D., and Ye, Y., *Linear and Nonlinear Programming*, Springer Science + Business Media, LLC, 2008.
- [25] *Gurobi Optimizer Reference Manual*, Gurobi Optimization LLC, 2018.